Elizabeth Newton   **15.075 Applied Statistics**   **Spring 2003**

Notes on S-Plus

In this course, we will be working mostly at the command prompt, though functions can be run from the menus as well. To get to the command prompt open the commands window. When you install S-Plus you can specify that this window open automatically on start-up.

Objects in S+ are either functions or data. To see a dataset just type its name, for instance **x.** If **x** is a matrix, and you want to see, say, only the second column, type **x[,2]**. Or if you want to see the second row, type **x[2,]**

To assign a numeric value, say 1, to an object, for instance, **x**, type **x<-1** (instead of x=1).

To make a little vector of 1, 2 and 3 you could type **x<-c(1,2,3)** or **x<-1:3** or **x<-seq(1,3,1) .** c means "concatenate" It is a very useful function, but it means you don't want to name a data object c. To make a vector of 3 one's type **x<-rep(1,3).**

For information about any function type **?functionname**, for instance **?seq**. Or type **help(functionname)**.

To see the statements which make up a function, type the function name with no parentheses after it, for instance **sample**.

To run a function, type the function name with parentheses including any required arguments. For instance **sample(x).**

To start creating or edit a function of your own, called, for instance, myfunction, type **fix(myfunction)**. Put the arguments, if any, inside the parentheses and function statements inside the {}'s. To assign a default value to an argument, for instance, 10, say **x=10.** When you save and exit from the editor, if S-Plus finds any mistakes, it will complain. Type **fix()** rather than fix(myfunction), to return to where you were.

If a dataset has missing values, indicated by **NA,** most functions (for instance, **mean**) simply will return a value of NA. To get a mean of the non-missing values of **x**, type: **mean(x, na.rm=T)**. It is not one of the greatest features of S+, that there are many ways of telling functions to ignore missing values. To find out which way a particular function uses, simply type **?functionname**.

Data Objects (see chapter on data objects in *Programmers Guide*).

There are four basic types of data objects for organizing data and computational results. These are data frames, matrices, vectors, and lists. You can ask which type an object is. For instance, to find whether the object, x, is a data frame, type is.data.frame(x) which will return T or F. Similarly there are functions: is.vector, is.matrix, is.list, is.function.

Data Frames

In S-PLUS, the primary structure for storing two-dimensional data is the data frame. The data frame can contain columns of different modes. For example, in a two-column data frame, the first column can contain numeric data, while the second column can contain character data.

The example data frame kyphosis has 81 rows of data on 81 children who have had corrective spinal surgery. It has four columns, representing the variables Kyphosis, Age, Number, and Start. Kyphosis is a two-level factor telling whether a postoperative deformity (kyphosis) is "present" or "absent". The other three variables are numeric vectors. Age is the age of the child in months. Number is the number of vertebrae involved in the operation. Start is the beginning of the range of vertebrae involved in the operation.

Matrices

Matrices are similar to data frames, except that all elements of a matrix must contain data of the same mode. Commonly used modes are character, numeric, complex, and logical. Matrices can have both row and column names. Matrices may be used to store data, but the user is more likely to encounter them as the results of some computation.

The example matrix cereal.attitude gives the percentage of people agreeing with 11 statements, such as "Reasonably Priced", about 8 brands of cereals. It has 11 rows and 8 columns.

Vectors

A vector is an ordered set of elements having the same mode. Commonly used modes are character, numeric, complex, and logical. Like the rows and columns of matrices and data frames, the elements of a vector can have names. Each column in a data frame is a vector. Vectors also result from some computations.

The example vector ozone.median gives 41 median ozone readings taken over time.

Lists

Lists are collections of other objects. Their components can be data frames, matrices, vectors, other lists, functions, or any other S-PLUS objects. Lists are used to contain related data objects such as computational results from a linear regression fit. A data frame is a special type of list.

The example list evap has a component x which is a matrix of independent variables and a component y which is a vector of daily evaporation amounts.